УДК 681.3

ПОСТРОЕНИЕ ЛОГИЧЕСКИХ МОДЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ ДЕРЕВЬЕВ РЕШЕНИЙ

О.Г. Берестнева, Е.А. Муратова

Томский политехнический университет
E-mail: helmu@mail.ru

Рассматриваются методы выявления скрытых закономерностей в виде деревьев решений. Предложена информационная технология, позволяющая выявлять присущие исследуемой предметной области устойчивые закономерности в виде логических правил (дихотомических деревьев решений). Эффективность разработанной технологии продемонстрирована на примере решения задачи определения специфики когнитивной сферы лиц с различными типами интеллектуальной успешности.

Введение

Статья продолжает цикл статей [1, 2], посвященных проблеме формирования базы знаний для интеллектуальных систем. В работах [1–7] нами были рассмотрены методы, позволяющие выявить присущие исследуемой предметной области устойчивые закономерности на основе имеющихся данных с привлечением или без привлечения экспертов. Показано, что результаты решения одной и той же диагностической задачи различными методами иногда приводят к противоречивым выводам. В [1] предложен подход, обеспечивающий совместное использование полученных решений с целью повышения качества распознавания, классификации и прогнозирования при использовании минимального количества диагностических прецедентов; разработана технология конструирования диагностических решений в виде логических правил. В данной работе подробно рассматривается один из методов формирования знаний – технология построения логических моделей в виде деревьев решений.

Постановка задачи и описание методов решения

Задача построения логических правил не нова [1, 2, 7, 8–10]. Однако, до настоящего времени, остается актуальной задача разработки технологии совместного использования разных диагностических решений, что будет особенно ценным, например, для выборок малой размерности, характерных для социологии, психологии и психодиагностики. Исходным материалом при решении задач анализа данных является набор объектов предметной области, представленный характеризующими его признаками, которые измеряются в шкалах разного типа.

Деревья решений являются достаточно распространенным в настоящее время подходом к выявлению и визуализации логических закономерностей в данных [11–13]. В работе используются дихотомические деревья, когда из вершины выходит только две ветви. Каждому узлу сопоставлен некоторый признак, а ветвям – либо конкретные значения для качественных признаков, либо области значений для количественных признаков.

Дерево решений позволяет построить модель зависимости множества исходов от множества характеристических признаков.

При построении дерева решений должно соблюдаться требование непротиворечивости – на пути, ведущем из корня в лист, не должно быть взаимоисключающих значений. Дерево решений может быть переведено в набор логических высказываний. Каждое высказывание получается при прохождении пути из корневой вершины в лист, и представляет собой логическую закономерность исследуемого явления.

Качество дерева характеризуют два основных показателя: точность и сложность дерева. Точность дерева показывает насколько хорошо разделены объекты разных классов. В качестве показателя сложности дерева выступают такие характеристики как: число листьев дерева, число его внутренних вершин, максимальная длина пути из корня в конечную вершину и др. Показатели сложности и точности взаимосвязаны: чем сложнее дерево, тем оно, как правило, точнее.

В наших исследованиях для построения модели дерева решений был использован алгоритм [12], работа которого начинается с того, что на вход поступает некоторое количество обучающих примеров (объектов). Каждый объект описывается набором характеристических признаков (в дальнейшем также разделяющие признаки) и классифицирующим признаком, который задает принадлежность к одному из диагностических классов. Корню дерева соответствует самый информативный характеристический признак. Далее, в вершинах располагаются признаки в порядке уменьшения значений прироста информативности. В качестве меры информативности узла используется энтропия.

Рассмотрим этот процесс более подробно.

Пусть имеется множество T объектов, разделенных по значениям классифицирующего признака на полные непересекающиеся классы C_1, C_2, ..., C_k (классифицирующий признак может принимать k возможных значений), тогда информация, необходимая для идентификации класса, есть

$$\mathrm{Info}(T) = I(P),$$

где P – вероятность распределения классов (C_1, C_2, ..., C_k):

$$P = (p_1, p_2, ..., p_k) = \left(\frac{|C_1|}{|T|}, \frac{|C_2|}{|T|}, ..., \frac{|C_k|}{|T|}\right),$$

а $I(P)$ – энтропия, вычисляемая по формуле:

$$I(P) = -(p_1 \cdot \log_2(p_1) + p_2 \cdot \log_2(p_2) + ... + p_k \cdot \log_2(p_k)).$$

Информация, необходимая для идентификации класса при условии, что нам известно значение разделяющего (характеристического) признака X, считается как:

$$\text{Info}(X,T)=\sum_{i=1}^{m}\left(\frac{|T_i|}{|T|}\cdot\text{Info}(T_i)\right),$$

где T_i – одно из возможных значений разделяющего признака X, m – количество значений разделяющего признака, $\text{Info}(T_i)$ – информация для каждого значения разделяющего признака.

Тогда величина, характеризующая прирост информативности $\text{Gain}(X,T)$ может быть определена как:

$$\text{Gain}(X,T) = \text{Info}(T) - \text{Info}(X,T).$$

Прирост информативности представляет собой разницу между информацией, необходимой для идентификации класса и информацией, необходимой для идентификации класса при условии, что нам известно значение признака X. При использовании обучающей выборки с неполным набором информации вычисление коэффициента прироста признака производится только по признакам с определенными значениями.

Понятие "прирост информации" необходимо для ранжирования характеризующих признаков при построении дерева решений. Каждый новый узел, включаемый в дерево решений, располагается так, что он приносит наивысший прирост информативности из всех разделяющих признаков, еще не включенных в путь к корню.

При последующих ветвлениях может возникнуть ситуация, когда вероятностное распределение разделяющего признака D представляет собой (1,0). Тогда $\text{Info}(D,T) = 0$ и $\text{Gain}(D,T)$ максимален. Чтобы это компенсировать используется вместо коэффициента Gain следующий коэффициент:

$$\text{GainRatio}(D,T)=\frac{\text{Gain}(D,T)}{\text{SplitInfo}(D,T)},$$

где $\text{SplitInfo}(D,T)=I\left(\frac{|T_1|}{|T|},\frac{|T_2|}{|T|},\ldots,\frac{|T_m|}{|T|}\right)$, а $\{T_1, T_2, \ldots, T_m\}$ – подмножества T_i, порождаемые делением множества объектов в соответствии со значениями признака D.

Если качественный признак, то при вычислении коэффициента прироста информации используется каждое значение. Количественный признак требует предварительных разбиений на некоторые градации или интервалы. Рассмотрим как это происходит.

Пусть признак C_i – количественный. Возможные значения признака сортируются в порядке возрастания: $A_1, A_2, \ldots, A_m$, затем для каждой величины $A_j (j=1, 2, \ldots, m)$, записи разделяются на те, которые имеют значения до A_j включительно и те, которые имеют значения большие A_j. Для каждого из полученных подмножеств вычисляется прирост или коэффициент прироста информации. В итоге выбирается деление на подмножества с максимальным коэффициентом прироста. Полученное пороговое значение подлежит проверке или уточнению в ходе дальнейших исследований.

Вершина относится к бесперспективным для последующего ветвления в случае, если объекты обучающей выборки для данной вершины однородны (принадлежат одному диагностическому классу), или число объектов достаточно мало (порог на число наблюдений задается в качестве входного ограничивающего параметра алгоритма).

Усечение дерева решений производится путем замещения целого поддерева узловым листом. Замещение имеет место только в том случае, если ожидаемый показатель ошибки в поддереве больше, чем в одиночном листе.

На первом шаге вычисляется количество ошибок $E_0(t)$ в поддереве с корнем в вершине t:

$$E_0(t)=\sum_{l=1}^{L}\sum_{\substack{\omega=1\\ \omega\neq Y(t)}}^{K}U_l^{\omega},$$

где L – количество вершин, на которые разделилась вершина t, K – количество диагностических классов, которые соответствуют вершине t, $\omega\neq Y(l)$ – дополнительное условие, соответствующее тому, что не рассматриваются решения (классы), которые были присвоены предыдущим вершинам, U_l^{ω} – число объектов класса ω, которые соответствуют l-ой вершине.

На втором шаге подсчитывается количество ошибок $E_1(t)$ которое будет допущено, если поддерево будет преобразовано в лист. Затем вычисляется выигрыш $G(t) = E_0(t)-E_1(t)$. Вершина, имеющая большое значение выигрыша, подлежит усечению.

Технология построения деревьев решений в системе See5

В настоящее время известны несколько десятков компьютерных программ для построения деревьев решений. Одними из самых популярных в мире сейчас являются программные системы CART (предназначенные для решения задач распознавания образов и регрессионного анализа), C4.5 или модернизированный вариант этой системы See5 (для решения задач распознавания) [10]. Система See5/C5.0 компании RuleQuest предназначена для анализа больших баз данных, содержащих до сотни тысяч записей и до сотни числовых или номинальных полей. Результат работы See5 выражается в виде деревьев решений и множества if-then-правил. Задача See5 состоит в предсказании диагностического класса какого-либо объекта по значениям его признаков. Остановимся более подробно на основных этапах обработки и анализа данных в этой системе.

Первый этап. Подготовка данных к анализу. Система See5 требует задание двух обязательных файлов: первый с перечислением имен разделяющих признаков и указанием классификационного признака (файл с расширением "*.names") и второй с данными (файл с расширением "*.data"), где по строкам располагаются объекты, а по столбцам

```
iqt <= 107 (108.5): 0 (45.1/7.5)  (лист)
iqt >= 131 (108.5):               (вершина)
:...uit2 >= 49 (39): 1 (4.4/1.1)
 uit2 <= 29 (39):
 :...iq <= 110 (111): 0 (22.5/3.7)
 iq >= 125 (111):
 :...nk2 >= 0.71 (0.355): 1 (31.2/10.5)
 nk2 <= 0.3 (0.355):
 :...iq <= 116 (117): 0 (7.5/0.7)
 iq >= 118 (117):
 :...time <= 52 (54): 1 (4/0.7)
 time >= 56 (54): 0 (12.3/3.7)
```

Рис. 1. *Иерархическая структура дерева решений*

признаки, причем в том порядке, в котором они заданы в файле названий. Кроме того, могут быть сформированы необязательные файлы с контрольной выборкой, где предсказываемая характеристика известна (файл с расширением "*.test") или неизвестна (файл с расширением "*.cases"). Все создаваемые файлы, предназначенные для решения одной задачи анализа должны иметь одинаковое имя. Файлы могут быть сформированы в любом текстовом редакторе.

Второй этап. Задание начальных параметров и построение дерева решений. В качестве параметров (через пункт меню Construct Classifier) могут задаваться следующие:

– возможность перевода деревьев решений в коллекцию логических правил;
– ранжирование правил по уровню значимости – от наибольшего к наименьшему.
– построение леса решений методом случайных подвыборок;
– объединение отдельных значений в подмножества для уменьшения ветвления дерева (по умолчанию каждому значению соответствует одна исходящая ветка);
– опция для деления выборки на контрольную и обучающую (задается размер обучающей выборки в процентах);
– L-кратная перекрестная проверка получаемых решений;
– задание нечетких границ подмножеств значений;
– задание числа вершин (в процентах от исходного дерева), которое должно остаться после усечения;
– минимальное количество объектов, соответствующих выделенному листу.
– Результатом выполнения второго этапа являются сконструированные деревья решений.

Третий этап. Анализ полученных правил. Сконструированные деревья решений, удовлетворяющие заданным параметрам, на представляются в виде иерархической структуры (рис. 1), и в виде логических правил (рис. 2).

```
Extracted rules:

Rule 2/1: (34.2, lift 1.5)
 iqt <= 107
 -> class 0 [0.972]

Rule 2/2: (50.5/16.6, lift 1.0)
 iqt > 107
 uit2 <= 33
 -> class 0 [0.666]

Rule 2/3: (17.9/5.1, lift 2.0)
 iq > 110
 iqt > 107
 nk2 > 0.35
 -> class 1 [0.692]

Rule 2/4: (3.8/0.8, lift 2.0)
 uit2 > 33
 -> class 1 [0.685]

Rule 2/5: (38.6/19.1, lift 1.5)
 iq > 110
 iqt > 107
 -> class 1 [0.506]
```

Рис. 2. *Логические правила, соответствующие дереву решений на рис. 1*

Каждому листу дерева приписывается некоторое число – значение прогнозируемой характеристики (номер диагностируемого класса), после которой в скобках указываются параметры или (n), или (n/m). Число n обозначает количество объектов, относящихся к данному классу и второе число m (если такое появляется) – количество ошибочных классификаций для данного листа. Для каждого дерева решений выводятся характеристики: число листьев дерева и точность классификации (в данном случае это коэффициент ошибок); приводятся результаты правильной и ошибочной классификации для каждого диагностического класса и время построения дерева решения.

Каждое дерево решений представляется также в виде множества логических правил (если задана соответствующая опция). Например, для приведенного на рис. 1 дерева система выделила 5 правил, представленных на рис. 2. Каждое правило, выводимое системой, характеризуется величинами (n/m, lift x): n – количество объектов, соответствующих данному правилу; m – количество объектов, не принадлежащих данному диагностическому классу (ошибочное распознавание); lift x – уровень доверия к построенному правилу.

Уровень доверия вычисляется по формуле

$$\text{lift}\, x = \frac{A}{f},$$

где A – точность правила, оцениваемая с помощью соотношения Лапласа

$$A = \frac{n - m + 1}{n + 2};$$

f – относительная частота прогнозируемого класса по всей обучающей выборке

$$f = \frac{N_1}{N},$$

где N_1 – количество объектов, соответствующих прогнозируемому правилом классу, N – объем обучающей выборки.

Четвертый этап. Построение леса деревьев решений.

Для того чтобы улучшить качество классификации, распознавания и прогнозирования, а также для получения устойчивых закономерностей (под устойчивостью автор понимает повторение результатов) исследуемого явления в системе See5 предусмотрена процедура построения леса деревьев решений. Деревья могут быть получены разными методами (или одним методом, но с различными параметрами работы), по разным выборкам. В выбранной системе реализован адаптивный метод, основная идея которого состоит в том, что для формирования деревьев решений используются различные части исходной обучающей выборки [13]. Вначале каждому объекту приписывается равная вероятность отбора в подвыборку, и по всей исходной выборке строится первое дерево решений. На следующих этапах вероятность отбора каждого объекта изменяется: неправильно классифицированные объекты получают приращение вероятности на заданную величину. Формируется следующая подвыборка с учетом новых вероятностей отбора, по которой строится другое дерево решений. Процедура продолжается до тех пор, пока не будет построено заданное исследователем количество деревьев.

Для построения коллективной классификации и прогнозирования используется метод голосования, т.е. объекту приписывается тот класс, которому отдает предпочтение большинство деревьев из набора.

Результатом для леса решений является общая точность классификации.

Применение логических моделей и деревьев решений при исследовании специфики когнитивного обеспечения интеллектуальной деятельности студентов

Изложенная выше технология построения деревьев решений была использована при проведении исследований в области психологии интеллекта, проводимых в рамках проекта РФФИ "Моделирование механизмов эффективной интеллектуальной самореализации субъекта".

Математическое моделирование эффективного функционирования интеллекта, как особой формы ментального опыта субъекта, является принципиально новым для современной когнитивной психологии. Огромный материал в области изучения интеллекта до сих пор не позволяет определить конкретные компоненты когнитивной сферы, которые способствуют человеку максимально продуктивно использовать свои возможности.

Особую актуальность имеет изучение когнитивно-стилевой организации интеллекта, которая, по мнению М.А. Холодной, проявляется в способности к непроизвольному контролю интеллектуальной деятельности, оказывая тем самым влияние на интеллектуальную продуктивность личности [14].

Задачей нашего исследования было выявление специфических структурных компонент интеллекта и когнитивных стилей, необходимых для успешной интеллектуальной самореализации на основе анализа показателей продуктивности интеллектуальной деятельности и когнитивно-стилевой организации лиц с высоким и сверхвысоким интеллектуальным потенциалом.

В целях выявления результативных и стилевых показателей интеллектуальной деятельности использовались следующие методики:

1. Интеллектуальная шкала Амтхауэра (измерение уровня общего интеллекта в виде коэффициента интеллекта – IQ);
2. Два субтеста из шкалы Амтхауэра, а именно субтест 2 "Определение общих признаков" (так называемый "вербальный" интеллект) и субтест 6 "Ряды чисел" ("технический" интеллект).
3. Стилевые методики, а именно: методика "Включенные фигуры" Уиткина, индивидуальный вариант (измерение когнитивного стиля полезависимость/поленезависимость); методика "Сравнение похожих рисунков" Кагана (измерение когнитивного стиля импульсивность/рефлективность); методика Струпа "Словесно-цветовая интерференция" (измерение когнитивного стиля ригидность/гибкость познавательного контроля).

В качестве испытуемых выступали успешно обучающиеся студенты и магистранты вузов г. Томска. Выборка состояла из 127 человек. При этом, 39 из них имели реальные достижения в интеллектуальной сфере деятельности (именные стипендии, гранты, публикации, участие в научных конференциях и т.д.). Далее в тексте для данной группы испытуемых используется название "успешные". Особенность выборки состояла в том, что абсолютно интеллектуально непродуктивные личности в состав испытуемых не входили, однако, в нашем исследовании группа остальных испытуемых (88 человек) названа группой "неуспешных" (с точки зрения наличия реальных интеллектуальных достижений).

Таблица 1. *Описание и обозначение используемых переменных*

Обозначение	Описание
uspex	Классификационный признак
sex	Пол (0 – мужской; 1 – женский)
iq	Показатель общего интеллектуального развития IQ
iqv	Способность к понятийной абстракции
idt	Способность к индуктивному мышлению
t1	Время чтения 1-ой карты в тесте Струпа
t2	Время чтения 2-ой карты в тесте Струпа
t3	Время чтения 3-ей карты в тесте Струпа
if	Интерференция сенсорно-перцептивных и вербальных функций (t3/t2)
int	Интеграция сенсорно-перцептивных и вербальных функций (t2–t1)
time	Время принятия решения в тесте Кагана
er	Рефлективность (количество ошибок в тесте Кагана)
uit1	Время выполнения 1-ой половины теста Уиткина
uit2	Время выполнения 2-ой половины теста Уиткина
nk2	Имплицитная обучаемость (тест Уиткина)

В табл. 1 приведены обозначения и описание, используемых в работе переменных. Классификационный признак **uspex** принимает значение 1, для группы "успешных" и значение 0 – для группы "неуспешных".

Задача решалась на базе пакета See5. В качестве исходных параметров задавались: построение леса решений, количество деревьев – 10; объединение отдельных значений в подмножества для уменьшения ветвления дерева; перевод деревьев в логические правила и нечеткие границы подмножеств значений.

Для каждого построенного дерева был сформирован свой набор правил (табл. 2). На рис. 1 приведен пример одного из таких деревьев, а на рис. 2 – логические правила для данного дерева, соответствующие этому дереву решений.

Среди построенного множества логических правил встречаются правила, которые представляют собой явные закономерности. Так, например, 1 и 5 правило для дерева № 2 (рис. 2) указывают на известный факт, что для успешной интеллектуальной самореализации прогностически важным фактором является общий уровень интеллекта больше 107.

Таблица 2. *Характеристики логических правил*

№ дерева	Размер дерева (количество листов)	Количество ошибок распознавания	Количество логических правил	Количество ошибок распознавания
1	10	17 (13,4 %)	8	19 (15,0 %)
2	4	30 (23,6 %)	4	21 (16,5 %)
3	7	31 (24,4 %)	5	28 (22,0 %)
4	9	23 (18,1 %)	9	9 (7,1 %)
5	9	21 (16,5 %)	7	13 (10,2 %)
6	14	25 (19,7 %)	9	23 (18,1 %)
7	15	19 (15,0 %)	10	9 (7,1 %)
8	11	23 (18,1 %)	6	16 (12,6 %)
9	10	31 (24,4 %)	7	27 (21,3 %)
10	9	21 (16,5 %)	8	15 (11,8 %)
Всего	–	21 (16,5 %)	64	7 (5,5 %)

В табл. 2 приведены основные характеристики полученных логических правил. Как видно из таблицы, количество ошибок при распознавании меньше, если использовать группу логических правил. Этот факт объясняется тем, что правила не являются точной копией одного из путей от главной вершины дерева к листу, а представляют собой отдельные логические высказывания.

Из общего списка правил были исключены правила, дублирующие ниже описанные зависимости (как правило, состоящие из одного высказывания) и не представляющие самостоятельного интереса. В конечном итоге было выделено несколько групп логических правил.

Первая группа правил охватывает большую часть испытуемых, и выводит на первый план показатели интеллектуальных способностей. Зависимость в выделенных подгруппах такова: если показатели вербальных и числовых способностей, а также общего интеллекта относительно низкие (в пределах 102–106), то имеет место "неуспешность".

Вторая группа правил подчеркивает роль пола в сочетании с различными стилевыми характеристиками (заметим, что эти правила охватывают малые по объему подгруппы от 10 до 12 испытуемых). Зависимости таковы: если у мужчин наблюдается максимальная поленезависимость при выполнении 2-ой половины теста Уиткина, высокая имплицитная обучаемость, очень замедленный темп принятия решений в тесте Кагана и гибкость познавательного контроля, то эти испытуемые "успешны". Фактически, эти правила доказывают важную роль стилевых свойств интеллекта, причем их роль наиболее ярко проявляется в мужской части выборки.

Третья группа правил характеризует частные – и, что характерно, – *разнонаправленные* связи между продуктивными и стилевыми свойствами. Зависимости таковы: "неуспешность" имеет место, во-первых, если низкие показатели интеллектуальных способностей и низкий уровень общего интеллекта сочетаются с ригидностью познавательного контроля (в виде высокой интерференции), и, во-вторых, если низкие показатели интеллектуальных способностей и низкий уровень общего интеллекта сочетаются с ярко выраженной поленезависимостью (высокой скоростью выполнения 2-ой половины теста Уиткина) и высокой имплицитной обучаемостью. Вторая часть данной зависимости оказалась довольно неожиданной, и ее объяснение на данном этапе исследования представляется затруднительной.

Четвертая группа (небольшая по количеству входящих в нее правил) дополняет частные связи. Зависимости таковы: "неуспешность" наблюдается в случаях выраженности таких стилевых свойств, как импульсивность (в виде большого количества ошибок), полезависимость (в виде относительно медленной скорости выполнения 1-ой половины теста Уиткина), дезинтеграции словесно-речевых и сенсорно-перцептивных функций. Эти правила характерны в основном для женщин.

Пятая группа правил имеет особую значимость, так как является доказательством наличия *эффекта крайних значений* применительно к стилевым качествам интеллекта. Зависимости таковы: при максимальной выраженности разных стилевых свойств наблюдается "неуспешность" реальной интеллектуальной деятельности. Например, при максимальной поленезависимости (при выполнении 2-ой половины теста Уиткина) (средние значения 5 с), максимальной быстроте принятия решений (в среднем 21,5 с), максимально большом количестве ошибок (более 2). Эти факты подтверждают ранее сформулированную закономерность, согласно которой у испытуемых, занимающих крайние позиции на стилевой оси, снижается эффективность интеллектуальной деятельности [14–16].

Таким образом, проведенное исследование доказало наличие определенного симптомокомлекса интеллектуальных качеств, которые благоприятствуют реальным интеллектуальным достижениям человека в профессионально ориентированных видах научно-технической деятельности (высокий уровень развития понятийных и числовых способностей, а также сформированность мобильного поленезависимого, рефлективного, гибкого стилей переработки информации).

Заключение

Использование представленной в статье технологии конструирования диагностических решений на основе логических моделей в виде деревьев решений позволяет эффективно решать задачи выявления скрытых логических закономерностей в условиях разнотипных данных.

Эффективность предложенного подхода была продемонстрирована на примере решения задачи определения специфики межпроцессуальных взаимосвязей когнитивных процессов и исследования специфики организации ментального опыта субъектов с высоким уровнем интеллектуального развития. Несмотря на имеющиеся ограничения рассмотренных методов, авторам удалось выявить важные устойчивые закономерности. Таким образом, предлагаемая технология построения логических моделей на основе деревьев решений представляет практическую ценность для задач получения знаний для экспертных систем при использовании стратегии формирования знаний (т.е. выявления скрытых закономерностей с применением специального математического аппарата и программных средств).

Работа выполнена при финансовой поддержке РФФИ, проект № 03-06-80128

СПИСОК ЛИТЕРАТУРЫ

1. Муратова Е.А., Берестнева О.Г. Выявление скрытых закономерностей в социально-психологических исследованиях // Известия Томского политехнического университета. – 2003. – Т. 306. – № 3. – С. 97–102.
2. Муратова Е.А., Берестнева О.Г., Янковская А.Е. Анализ структуры многомерных данных методом локальной геометрии // Известия Томского политехнического университета. – 2003. – Т. 306. – № 3. – С. 19–23.
3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – 608 с.
4. Александров Е.А. Основы теории эвристических решений. – М., 1975. – 254 с.
5. Аметов Р.В., Берестнева О.Г., Муратова Е.А., Янковская А.Е. Технология конструирования диагностических решений в слабоструктурируемых проблемных областях // Интеллектуальные системы (IEEE AIS'03)" и "Интеллектуальные САПР" (CAD-2003): Труды Международных научно-техн. конференций. – М.: Физматлит, 2003. – Т. 1. – С. 267–272.
6. Анализ состояния и тенденции развития информатики. Проблемы создания экспертных систем // Исследовательский отчет. Под ред. С.А. Николова. – София: Интерпрограмма, 1988. – 151 с.
7. Анастази А., Урбина С. Психологическое тестирование. – СПб.: Питер, 2001. – 688 с.
8. Берестнева О.Г., Иванов В.Т., Иванкина Л.И., Шаропин К.А., Муратова Е.А. Информационная система мониторинга здоровья студентов // Вестник Томского государственного университета. – 2002. – № 1 (II). – С. 196–201.
9. Берестнева О.Г., Кострикина И.С., Муратова Е.А. Применение современных информационных технологий в задачах психологии интеллекта // Интеллектуальные системы (IEEE AIS'03)" и "Интеллектуальные САПР" (CAD-2003): Труды Международных научно-техн. конференций. – М.: Физматлит, 2003. – Т. 2. – С. 236–240.
10. Берестнева О.Г., Муратова Е.А. Компьютерные технологии в психологическом эксперименте // Компьютерные технологии в науке, производстве, социологических и психологических процессах: Матер. III Междунар. научно-практ. конф. – Новочеркасск: ООО НПО "Темп", 2002. – С. 23–25.
11. Берестнева О.Г., Муратова Е.А., Кострикина И.С. Извлечение знаний в задачах психологии интеллекта с использованием системы WizWhy // Математика. Компьютеры, образование: Тез. Междунар. конф. – Пущено, 20–25 января 2003. – М.: Изд-во "РиХД", 2003. – С. 13.
12. Берестнева О.Г., Муратова Е.А., Кострикина И.С. Компьютерное моделирование специфики развития познавательных способностей // Компьютерное моделирование 2003: Тр. Междунар. научно-техн. конф. – СПб.: Нестор, 2003. – С. 396–398.
13. Берестнева О.Г., Муратова Е.А., Янковская А.Е. Эффективный алгоритм адаптивного кодирования разнотипной информации // Искусственный интеллект в XXI веке: Труды Междунар. конгр. – Т. 1. – М.: Физматлит, 2001. – С. 155–166.
14. Берестнева О.Г., Муратова ЕА., Ротов А.В., Гаврилов М.А. Математическое моделирование влияния психокоррекции избыточного веса на организм человека // Актуальные проблемы информатики: Сб. трудов VI Междунар. научной конф. – Минск, 1998. – С. 250–256.
15. Богомолов В.П. и др. Программная система распознавания Лорег: алгоритмы распознавания, основанные на голосовании по системам логических закономерностей. – М.: ВЦ РАН, 1998.
16. Бонгард М.М. Проблема узнавания. – М.: Наука, 1967. – 220 с.
17. Дюк В., Самойленко А. Data Mining: учебный курс. – СПб: Питер, 2001. – 368 с.

18. Осипов Г.С. Приобретение знаний интеллектуальными системами: Основы теории и технологии. – М.: Физматлит, 1997. – 112 с.
19. Холодная М.А. Психологические механизмы интеллектуальной одаренности // Вопросы психологии. – 1993. – № 1. – С. 32–39.
20. Холодная М.А., Кострикина И.С., Берестнева О.Г. Проблемы продуктивной реализации интеллектуального потенциала личности // Вестник Томского государственного педагогического университета. – 2002. – Вып 3 (31). – С. 45–50.
21. Quinlan J.R. Induction of Decision Trees // Machine Learning. – 198. – № 1. – P. 1–81.
22. User's Guide, WizWhy Version 2. – WizSoft Inc. – http://www.wizsoft.com.
23. Vtogoff P.E. Incremental Induction of Decision Trees // Machine Learning. – 1989. – № 4. – P. 161–186.

УДК 622.692.12